Quantifying the subjective: Psychophysics and the geometry of color

Citation for published version:
Isaac, AMC 2013, 'Quantifying the subjective: Psychophysics and the geometry of color', *Philosophical Psychology*, vol. 26, no. 2, pp. 207-233. https://doi.org/10.1080/09515089.2012.660139

Digital Object Identifier (DOI):
10.1080/09515089.2012.660139

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Philosophical Psychology

Quantifying the Subjective: Psychophysics and the Geometry of Color

Alistair M. C. Isaac

October 26, 2011

How can we develop a scientific theory of experience? There seems to be an insurmountable problem here. Experience is inherently *subjective*—only I have access to my experience—yet science demands *objectivity*, in the sense that any results must be accessible and repeatable by others. The "what it's like" of a conscious experience cannot be detached from the subject who experiences it, so how can it be made accessible and repeatable to a scientific community? How can we have an objective theory of something which appears inherently subjective?

The essential insight for answering this question is that by intra-subjective comparisons of experiences, we can produce evidence for an objective theory of the *relationship* between them, if not of their intrinsic natures. In practice, this general approach has been implemented since at least the ancient Greeks, but its refinement into a systematic method is usually dated to the pioneering work of Ernst Weber and Gustav Fechner in developing the field of psychophysics (Section 1). However, the procedures they proposed assume that sensations can be linearly ordered. The rest of this paper will investigate the development of psychophysical methods for a domain which cannot be characterized by a simple linear order, the experience of color. In Section 2 we will explore some of the difficulties for ordering colors and how these inspired a new geometrical paradigm at the hands of Bernhard Riemann. Riemann's work explored the idea that the structure of subjective experience cannot be determined *a priori*, but must be measured. The first to actually make such measurements, and produce a metrically significant model of color space, was Helmholtz (Section 3).

Section 4 examines the one serious attempt to apply Weber's method directly to color experience, the difficulties encountered, and the alternative techniques which emerged in its wake. Special attention will be paid to the status of evidence for similarities in the experience of color across individuals. Section 5 assesses the evidence for objective claims about the experience of color: what can we say with certainty about the color experience of others? Is there any sense in which I can break through the apparent subjectivity and assert that your experience of color and mine are *the same*?

1 How to Study Subjective Experience

How can I measure qualitative experience, something which is inherently *subjective*? Although the "what it's like" of experience cannot be directly investigated experimentally, *comparisons* between a subject's experiences can be investigated through repeatable experiments. In particular, we can present subjects with stimuli and ask them to report upon how they differ. For example, I may not be able to characterize scientifically what it's like for you to experience the brightness of a light, but I can present you with two lights and ask which one appears brighter to you.

Although such comparisons have long been made in practice, the codification of this insight into a well defined experimental method can be dated to Gustav Fechner (1801–1887) and his *Elements of Psychophysics*. Fechner recognized that, in the absence of a means to directly measure sensation, one could let easily measured properties of the stimuli stand as a proxy for degrees of sensation: "we measure here not the sensation itself, but only the stimuli or stimulus differences that produce equal sensations or equal differences between sensations" (Fechner, 1860a, p. 46). This procedure assumes that intensity of sensation is a function of stimulus strength. By varying the independent variable (stimulus strength), we can learn about the relative value of the dependent variable (intensity of sensation):

> Insofar as sensitivity is a variable, we should not seek for a constant as its measure. We may, however, look for (1) its limits and (2) its mean values; we may also investigate (3) how its variations depend on conditions; finally we may seek (4) lawful relations that remain constant during variation; the last are the most important. (Fechner, 1860a, p. 45)

Fechner collected and elucidated several distinct experimental methods for investigating these questions. He called this field "psychophysics" since it investigates the systematic relationship between physical stimuli and psychological experience.

The most fundamental concept for understanding the methods of early psychophysics is the *just noticeable difference.* Fechner's discussion of the just noticeable difference and its role in experimentation is heavily influenced by the work of his mentor and collaborator, Ernst Heinrich Weber (1795–1878). Weber asked subjects to compare the intensities of sensations caused by two stimuli. By holding one stimulus fixed and systematically varying some parameter (e.g. weight, temperature, color) of the second stimulus, he could discover the smallest amount of physical (and therefore easily measurable) difference between stimuli required to produce a noticeable difference between sensations. Other methods discussed by Fechner, such as the method of right and wrong cases and the method of average error, share with the method of just noticeable differences two methodological constraints: 1. a single physical parameter is varied, and 2. subjects make comparisons between two stimuli.

What kind of conclusions can be drawn from psychophysical experiments? Weber discovered through experimentation that our experience of the weight of an object is actually calculated from at least three distinct inputs: pressure, muscle activity, and temperature. As a consequence, Weber argued that the supposed "sense of touch" is really a collection of distinct sensory faculties, each of which contributes to the experience of an object's heaviness. Although we usually do not attend to these different aspects of our experience of weight, Weber was able to isolate each of them by controlling for the effects of the others. Let's look at two examples in order to see how this works (see Weber, 1846, for details).

Experiment one. Weber was interested in the question of whether or not information from muscle activity (as when, say, one hefts an object) contributed to judgments of weight. In order to experimentally test this question, he devised a single set of stimuli and two conditions. The stimuli were just a set of small boxes, all identical in appearance and size, yet differing slightly in weight. In the first condition, subjects were asked to lay their forearms on a table (this was to prevent muscle activity from contributing to their judgments) and boxes were placed successively in their hands until the subject could detect the difference in weight. In this condition, only the sensation of pressure contributes to the experience of heaviness. Weber was able to measure the size of a difference in the weight of stimuli necessary to produce a just noticeable difference in the subject's perception of heaviness. In the second condition, Weber again offered the subject boxes differing in weight until the subject was able to notice the difference. This time, however, the subject was standing up and able to heft the boxes in each hand before judging. Weber again measured the difference in stimulus weight required to produce a just noticeable difference in the subject's perception of heaviness. What Weber discovered was that the difference in weights required to produce a perceptual difference in heaviness was much smaller in the second condition, when subjects were allowed to use their muscles, than in the first condition. He concluded that information from muscle activity indeed contributes to judgments of weight.

Experiment two. Weber wondered whether sensibilities other than pressure and muscle activity contribute to judgments of weight. Weber tested whether the sensation of temperature contributes to such judgments by placing coins of different temperatures on a subject's forehead (i.e. out of sight of the subject). Weber discovered that cold coins were judged to be much heavier than warm ones. For example, if a single silver Thaler cooled to near freezing and a stacked pair of silver Thaler at room temperature are placed on a subject's head in alternation until the subject can form a judgment about their respective heaviness, the subject will judge them to be equally heavy, or even that the single cold Thaler is heavier than the two stacked warm Thaler. Weber concluded that sensations of temperature indeed contribute to judgments about weight.

These two experiments support *qualitative* conclusions about how the subjective experience of heaviness is calculated by our perceptual system. In order to develop a *quantitative* theory of subjective experience, we merely need to perform Weber's experiments with a number of different comparison weights. This will reveal any systematic relationship between changes in weight and changes in the experience of heaviness. If such exists, it would constitute an instance of Fechner's "lawful relations that remain constant during variation."

Weber discovered not only that such systematic relationships do exist, but that they seem to follow the same basic pattern in many different sensory modalities.[1] In particular, the change in a stimulus required to produce a just noticeable difference in the subject's experience appears to stand in a constant proportion to the comparison stimulus, i.e.

$$\frac{\Delta S}{S} = k,$$

where ΔS is the change in stimulus strength, S is the absolute value of the comparison stimulus, and k is a constant which differs for each sensory modality.

As a simple illustration of this principle, consider the change in the sensation of brightness caused by the lighting of a candle. In a dark room, a quite dramatic change occurs, in a well lit room, hardly any. Here, the darkness or lightness of the room is the standard of comparison S and the change in stimulus ΔS is measured by the number of candles lit. The fact that a greater change in stimulus (i.e. more candles) is needed to produce a noticeable difference in sensation when S is large (the well lit room) is a direct consequence of Weber's principle. Weber was able to quantify sensitivity to differences by measuring k in different experimental conditions. For example, in experiment one, condition one, Weber measured $k = 1/30$. In condition two, when muscle activity was allowed to contribute to assessments of heaviness, he measured $k = 1/40$ (Weber, 1846, p. 220).

Fechner dubbed Weber's discovery "Weber's Law" and discussed the empirical evidence, both for and against it, in detail. He was well aware that it holds only approximately for many sensory modalities and fails to hold entirely in others (Fechner, 1860a, Ch. IX). Nevertheless, he recognized that Weber's Law had value as a precise hypothesis against which quantitative empirical data could be assessed.

In order to further this project, Fechner supplemented Weber's Law with a powerful, but controversial, assumption. He assumed that just noticeable differences measured against different comparison stimuli correspond to equal changes in the intensity of experience. This allowed Fechner to interpret Weber's Law as describing a relationship between changes in stimulus strength and changes in sensation:

$$\Delta I = k \frac{\Delta S}{S},$$

where ΔI is the change in intensity of sensation. By taking the integral, he transformed this into a formula characterizing the relationship between experience and stimulus.

$$I = k \log \frac{S}{S_0}$$

S_0 here is the absolute threshold, the smallest stimulus value at which a sensation is noticed. Although later dubbed "Fechner's Law" by his followers, Fechner himself called this a "measurement formula." Since it "permits the amount of sensation to be calculated from the relative amounts of the fundamental stimulus," it can provide "a measurement of sensation" (Fechner, 1860b, p. 75).

Although Fechner's mathematics is sound, it is important to note that his assumption that all increments ΔI are equal in size represents an arbitrary conventional choice. It is not an hypothesis which can be empirically tested, merely a stipulation which allows Fechner to use stimulus size as a proxy for measuring sensation. However tendentious, this assumption opens the door for the mathematical analysis of psychophysical data and generates precise predictions about just noticeable differences which can be empirically tested.

The types of sensation examined by Fechner and Weber all vary along a single qualitative dimension. Sensations of loudness, brightness, pain, etc. all vary only in intensity, consequently, we can employ Fechner's methods in order to measure the relevant constant k, and allow discrepancies with Weber's Law to direct us toward a more nuanced mathematical analysis. In order to do the same for the sensation of color, however, more is required.

2 What's So Difficult about Measuring Color?

A presupposition behind Weber-style investigation of the experience of heaviness, brightness, etc. is that sensation in these domains has roughly the structure of a linear order. For any two sensations of loudness a and b, where $i(x)$ represents the intensity of sensation x, either $i(a) < i(b)$, $i(a) > i(b)$, or $i(a) = i(b)$.

This feature does not hold for our sensations of color. Consider a rich orange sensation o and a pale yellow sensation y. The orange sensation is more red, or at least closer to red, than the yellow sensation, $o >_R y$. Yet the yellow sensation is closer to white, i.e. much lighter than, the orange sensation, $y >_L o$. There is no translation between these two orderings: lightness cannot be derived from degree of redness, nor vice versa. So it looks like we need at least two distinct orderings to characterize color experience, maybe more. How should we determine the right geometric structure here? This section traces the development of the idea that color experiences form a geometric space of varying curvature, the structure of which can be determined empirically.

A critical thread in the development of more nuanced ideas about color space begins with Kant. Unfortunately, Kant's own discussion of the experience of color was relatively limited. He characterized the sensation of color in terms of *intensive magnitudes.* These are magnitudes because they have a greater or lesser degree, but they are "intensive" because they vary in intensity (in contrast to the "extensive" magnitudes of space and time).[2] There is one such scale of intensive magnitude associated with each possible color. "Every color, e.g. red, has a degree, which, however small it may be, is never the smallest, and it is the same with warmth, with the moment of gravity, etc." (A169, B211). Although the particular quality of a sensation is purely *a posteriori*, this structure, of being able to take on various degrees of value, from zero upwards in intensity, is deduced by Kant *a priori*:

> All sensations are thus, as such, given only *a posteriori*, but their property of having a degree can be cognized *a priori*. It is remarkable that we can cognize *a priori* of all magnitudes in general only a single quality, namely continuity, but that in all quality (the real of appearances) we can cognize *a priori* nothing more than their intensive quantity, namely that they have a degree, and everything else is left to experience. (A176, B218)

Kant proposes to determine the structural features of color perception *a priori*, and concludes that the experience of any particular color may vary continuously in degree from a zero point. Kant leaves out, however, any discussion of how different colors relate to each other.

Johann Friedrich Herbart (1776–1841) developed Kant's ideas in two important directions. First, he argued that the structure of color experience cannot be determined *a priori*, but must be discovered *empirically* (c.f. Lenoir, 2006, pp. 148*ff*). Second, he supplemented Kant's analysis of intensities with a discussion of the role of *similarities* between color experiences. Herbart's ideas on color were shaped during his time at Göttingen, where he studied from 1801 to 1809. After a professorship at Königsberg (the same chair held by Kant), he returned to Göttingen in 1833 where he remained until his death.

At Göttingen, Herbart was exposed to the ideas of Tobias Mayer (1723–1762). In the eighteenth century, the idea that colors should be arranged in a two dimensional space (as opposed to merely ordered along a line) was still young, with the first significant proposal being Newton's color circle (first edition of the *Opticks*, 1704, see discussion in Section 3). Mayer represents another important benchmark. He arranged hues into an equilateral triangle, with yellow, red, and blue at the vertices, and points in the interior representing mixtures of these three proportional to their distances from them. This base triangle was supplemented by a stack of gradually diminishing lighter versions in one direction and darker in the other, forming one of the first three dimensional color solids, a double tetrahedron (Kuehni and Schwarz, 2008, pp. 72–3).

Mayer was a professor at the University of Göttingen from 1751, and his followers continued to lecture on his system there for many decades after his death. His system influenced Thomas Young,[3] who studied at Göttingen in 1795 and 1796, and it influenced Herbart, who realized the challenge it posed to Kant's analysis of color experience.

In order to combine Kant's insight into intensive magnitudes with Mayer's analysis of color space, Herbart needed a unified framework for analyzing both the degree of a sensation and the similarity between distinct sensations. To solve this problem, he introduced the concept of a *series of transitions*. By following series of sensations, we can learn the structure of possibilities in a sensory domain. Furthermore, the following of a series is an empirical endeavor, and so the structure of color perception which for Kant was discovered *a priori* becomes, for Herbart, discoverable *a posteriori*. After comparing the type of series one can follow through space with those one can follow through time, number, and "degree or intensive magnitude" (the Kantian magnitudes), he remarks:

> Less distinct, but nevertheless indispensable, is the series produced by the putting together of sensations of the same kind according to the possibility of transition from one to another. From this we have the tone series .... Similar to it would be the color surface between the three primary colors, yellow, red, and blue, if we knew certainly whether all the colors were connected with the grades of difference between light and dark (perhaps we should say black and white), and could be traced back to those three; or whether the color realm does not rather require a third dimension. (Herbart, 1834, pp. 58–9)

Transitions across similar sensations motivate organizing them into a richer structure than can be found by just considering degrees of intensity. In the case of sound, the series of possible tones (organized by pitch) is an example; in the case of color, Mayer's triangular color surface. Notice that there is uncertainty here—waiting to be discovered *empirically* is whether the space of possible colors requires a third dimension.

Herbart outlines a general empirical procedure for determining series via judgments, illustrating it in the domain of color experience. (In the following A is a determinable, e.g. color, while a, b, c, etc. are its determinants, the specific values it can take, e.g. brick-red, vermilion, puce, etc.).

> A multitude of such judgments as A is a, A is b, A is c, A is d, etc.,

> . . . form a series; since the *a*, *b*, *c*, *d*, blend in different degrees according to their lesser or greater contrasts (e.g., the three judgments—this fruit is green, that yellow, a third yellowish green—blend in such a way as to bring with them the colors in their orders—green, yellowish green, and yellow; for between yellow and green the opposition is the strongest, consequently the blending the least). . . . The more the series of characteristics form and separate in this way, through the comparison of similarities, and in part of differences, so much the sooner will it be possible, by means of them, to determine the content of the complexes, or to approach the definitions of ideas. (Herbart, 1834, p. 147)

This passage was written before Fechner's codification of psychophysical methods, and fails to suggest any experimental procedure as rigorous as those employed by Weber. Nevertheless, the general strategy of using similarity judgments to explore the structure of a space of sensory experiences is essentially that of Weber and Fechner. Furthermore, Herbart, no doubt under the influence of Mayer, has emphasized the possibility that such similarity judgments may reveal an arbitrarily complex space, as a opposed to a simple linear arrangement.

Herbart's method, following a series of transitions, is limited in what it can reveal about a geometrical space. In particular, only *topological* structure (which points are connected to which other points) could possibly be revealed by following Herbart's suggestions. But a full understanding of the experience of color demands more than this; it demands also a knowledge of the *metrical* structure of color space, i.e. the *distances* between points. For example, there is some loose topological sense in which yellow and orange are "next to" each other, while yellow and blue are not. Much more interesting, however, would be to determine of a particular shade of orange whether it is closer to a particular shade of red or to a particular shade of yellow. Such a determination would require a meaningful *metric* for the space of possible color experiences.

In 1850 and 1851, Herbart's collected works were published and had a profound effect on Georg Friedrich Bernhard Riemann (1826–1866), then studying at Göttingen. He even developed his own theory of the relationship between mind and world along Herbartian lines (Ehm, 2010). In the lecture delivered on the occasion of his *Habilitation*, Riemann introduced a general framework for distinguishing topological from metrical structure in terms of *manifolds*, giving precision to ideas Herbart had discussed only loosely (Ferreirós, 1999, pp. 41–7)
.

Riemann begins with the concept of a manifold of specific "determinations"[4] and emphasizes the importance of Herbartian "transitions" for analyzing the connectedness of this structure, e.g. whether it is continuous or discrete.

> Notions of magnitude are only possible where there is an antecedent general concept which admits of different ways of determination. According as a continuous transition does or does not take place among these determinations, from one to another, they form a *continuous* or *discrete* manifold; the individual determinations are called points in the first case, in the last case elements, of the manifold.... so few and far between are the occasions for forming notions whose determinations make up a *continuous* manifold, that the only simple notions whose determinations form a multiply extended manifold are the positions of perceived objects and colours. (Riemann, 1868, p. 653)

Riemann makes a general comparison between physical space ("the positions of perceived objects") and the space of possible colors. In both cases, there are different determinations which may obtain (specific locations, specific colors), and in both cases, the transition through these points is continuous. These ideas are already present in Kant and Herbart, but Riemann supplements them with a discussion of metrical structure.

> Definite parts of a manifold, distinguished by a mark or by a boundary, are called quanta. Their comparison with regard to quantity is accomplished in the case of discrete magnitudes by counting, in the case of continuous magnitudes by measuring. Measuring consists in the superposition of the magnitudes to be compared; it therefore requires a means of transporting one magnitude as the standard for another. In the absence of this, two magnitudes can only be compared when one is a part of the other; in which case also we can only determine the more or less, and not the how much. (translation in Ferreirós, 1999, pp. 68–9)

The essential point here is that metric structure requires a standard for comparison. Furthermore, this standard must in some sense be mobile, so that it can be compared directly with the area to be measured, i.e. *superimposed upon it.* Without such a standard, distances cannot be measured.

Though the point may seem trivial when considering the measurement of, say, physical distances with a ruler, its significance for the psychophysical investigation of color is profound. In particular, what should play the role of a measurement standard for subjective experience? Furthermore, how could we possibly take a magnitude from one area of the domain and superimpose it on another? Weber's standard of comparison was the just noticeable difference. This standard seems to avoid the problem of superposition by appealing to the smallest possible distance. If smallest possible distances are always the same size (they are, after all, the "smallest"), we do not need to move one such distance around and hold it against the others to ensure they are "the same." As discussed above, however, the claim that just noticeable differences measure equal changes in experience cannot be given empirical meaning. It constitutes an arbitrary stipulation, which is only *pragmatically* justified by its power to generate precise quantitative predictions.

Nevertheless, additional assumptions are needed in order to measure the metric of color experience. Fechner's strategy only measures *local* distances, but how do these measurements fit together? We need some theory of the gross structure of a sensory domain if we are to use assessments of just noticeable differences as evidence for its metric. Without a theory of the gross structure of a sensory domain, we cannot define linear sequences through it. And without linear sequences, we cannot apply the method of just noticeable differences. In the case of color, we do not know antecedently if the space of possible color experiences is circular, triangular, or some other shape. What is its dimensionality? What are its symmetries? Without some strategy for first answering these questions, we cannot apply Fechner's methods to color experience. It was one of Helmholtz's great achievements to develop such a strategy, refining methods first pioneered by Newton to produce the first precise determination of linear sequences through color space.

3 Helmholtz Breaks the Symmetry

In the early 1850s, while Riemann studied Herbart at Göttingen, Hermann von Helmholtz (1821–1894) was studying him at Königsberg (Lenoir, 2006, pp. 148–60, c.f. Hatfield, 1990, p. 180). It was during this period that many of his crucial experiments involving the structure of color experience were performed. Although Helmholtz only came into contact with Riemann's ideas later in his career, his experiments forced him to distinguish metric from topological structure. He was the first to carry out Herbart's proposal and perform empirical measurements on the psychological distances between color experiences. Although these measurements failed to produce a metric in the strict sense, they succeeded in demonstrating that color space is *asymmetrical*.

Helmholtz's understanding of geometrical structure agrees completely with Herbart's empiricist take on Kant, as can be seen from the concluding remarks of Helmholtz (1876). There, he emphasizes that "[t]he axioms of geometry, taken by themselves out of all connection with mechanical propositions, represent no relations of real things." Nevertheless, "[a]s soon as certain principles of mechanics are conjoined with the axioms of geometry, we obtain a system of propositions which has real import, and which can be verified or overturned by empirical observations" (Helmholtz, 1876, p. 683). No less problematic than spatial structure is the structure of color experience, which Helmholtz acknowledges as the source of his interest in non-Euclidean geometries.

> Whilst Riemann entered upon this new field from the side of the most general and fundamental questions of analytical geometry, I myself arrived at similar conclusions, partly from seeking to represent in space the system of colours, involving the comparison of one threefold extended aggregate with another, and partly from inquiries on the origin of our ocular measure for distance in the field of vision. (Helmholtz, 1876, p. 675)

A "threefold extended aggregate" is just a three dimensional space. What were the two such spaces for representing color which Helmholtz was driven to compare? The answer to this question will emerge from an examination of Helmholtz's experiments on color mixtures.

Helmholtz's experiments on color mixtures were driven by challenges from Hermann Grassmann (1809–1877).[5] Grassmann had added mathematical precision to Newton's analysis of color space, which predicted that *every color has a complement*, i.e. another color with which it mixes to produce white.[6] An experiment by Helmholtz in the early 1850s seemed to demonstrate that this is not the case, but rather that *many colors have no complement at all*. Grassmann (1853) criticized Helmholtz's result with a formal proof, based on two very basic assumptions about the continuity of color space, that every color *must* have a complement. Grassmann's two assumptions were:

1. "[E]very impression of color of this kind may be analysed into three mathematically determinable elements—*the tint*, *the intensity of the color*, and *the intensity of the intermixed white*. The various tints form a continuous series of such a kind, that when we start from one color of this series and proceed forward, we finally arrive at the original color." (p. 435)

2. "[*I*]*f one of two mingling lights be continuously altered (whilst the other remains unchanged), the impression of the mixed light also is continuously changed.*" (p. 437, italics in the original)

Grassmann clearly identifies these as "assumptions." Not only are they natural assumptions to make, however, but they were shared by all serious researchers into color at the time, *including Helmholtz.*

In the next section of the paper, Grassmann attempts to use Helmholtz's own results to predict the complementary colors which Helmholtz failed to produce. For example, Helmholtz could not produce white by mixing red with any other color, but he found that red mixed with blue produced a pale violet, while red mixed with green produced a pale yellow. Grassmann argues that since there is a continuous transition between pale violet and pale yellow which passes through white, there must be a color lying on the continuous transition between blue and green which will produce this white in complement with red. Unfortunately, he cannot pinpoint the error in Helmholtz's experiment, concluding only "There is consequently a deficiency here" (Grassmann, 1853, p. 440). However, Helmholtz checked the continuous transition from blue to green for a complement to red and failed to find one: there must have been a hidden assumption in his experimental procedure which prevented him from finding the complement Grassmann's analysis predicts.

This hidden assumption, which Grassmann failed to emphasize, but Helmholtz eventually discovered, is that color space is *circular.* This *assumption* erroneously appears to follow as a *conclusion* if one confuses topological with metrical considerations. The error here dates back to Newton, whose system Grassmann merely formalized, yet was difficult to identify since its derivation was always left implicit. In fact, as long as one is only investigating the topological properties of color space, as in Grassmann's proof, there is no error: color space is topologically equivalent to a circle. Although the existence of complementary colors can be demonstrated via a purely topological proof, however, their determination demands the introduction of metric considerations. Ironically, Newton himself, having introduced the assumption of circularity, discovered the very same discrepancy between theory and experiment as Helmholtz.[7]

Proposition IV of Book One, Part II, of the *Opticks* states the basic properties of compound colors. Newton performed a series of experiments which first used a prism to separate sunlight into a spectrum, then isolated "homogeneal" (monochromatic) lights by selectively passing parts of the spectrum through slits in black paper. This allowed him to determine that

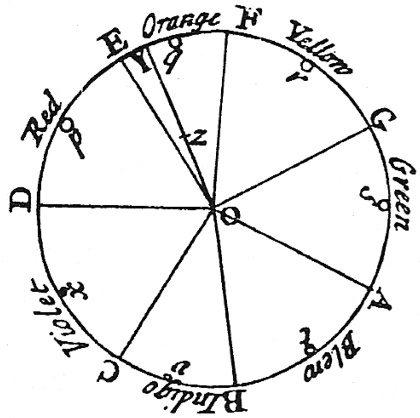

Figure 1: Newton's color circle (Newton, 1730, p. 155)

> ... Colours by how much they are more compounded by so much are they less full and intense, and by too much Composition they may be diluted and weaken'd till they cease, and the Mixture becomes white or grey. There may be also Colours produced by Composition, which are not fully like any of the Colours of homogeneal Light. (Newton, 1730, p. 132)

Intensity is not a technical term for Newton, but he seems to informally identify it with distance from white. Therefore, all the (maximally intense) homogeneous colors should be maximally distant from white. If one allows for a continuity of color across the red and violet ends of the spectrum, this motivates their organization into a perfect circle. Mixtures of these colors "dilute" them, this leaves them different in quality from homogeneous colors (different how? *closer to white*). Therefore, mixtures of homogeneous colors fall on the interior of the circle.

Proposition VI, Problem II, demonstrates how to use a color circle to calculate mixtures of colors. Newton proposes a strategy by analogy with the calculation of the center of gravity of a physical system. Pick points on the circle corresponding to the homogeneous colors being mixed. Assign them "mass" in proportion to the relative quantity (i.e. number of rays) of each color in the mixture. Finally, calculate the center of gravity of this system. Supposing, for example, that circles p, q, r, s, t, u, x represent the colors being mixed, with the size of the respective circles proportional to their "mass" in the mixture, then Z might be the center of gravity of this system, i.e. the color resulting from this mixture (see Figure 1).

Newton recognized imperfections in this system, acknowledging it as "accurate enough for practice, though not mathematically accurate" (Newton, 1730, p. 158). Furthermore, Newton acknowledges a crucial lacuna between the empirical predictions of his color circle and the effects he could produce experimentally:

> [I]f any two of the primary Colours which in the circle are opposite to one another be mixed in an equal proportion, the point Z shall fall upon the center O, and yet the Colour compounded of those two shall not be perfectly white, but some faint anonymous Colour. For I could never yet by mixing only two primary Colours produce a perfect white. (Newton, 1730, p. 156)

Anticipating Grassmann, Newton acknowledges the existence of complementary colors as a consequence of his model; yet, anticipating Helmholtz, he admits failure in experimentally producing the predicted effect!

Recognizing the severity of the challenge from Newton and Grassmann, Helmholtz resumed his experiments with a significantly more subtle apparatus. Before, he had assumed that complementary colors could be discovered by mixing *equal quantities* of homogeneous light. In his new setup, he abandoned that assumption.

Helmholtz isolated homogeneous light in the same manner as Newton had, by first passing sunlight through a prism, then allowing the resulting spectrum to fall upon a narrow slit which allowed only a single band, or color, of light to pass through. In order to mix two homogeneous lights, Helmholtz constructed a finely machined board with two adjustable slits. Both the vertical placement and the horizontal width of each slit could be precisely adjusted. Changing the vertical placement of a slit changed the wavelength of light it permitted, while changing the width of a slit changed the *quantity* of that light. By focusing the streams of light passing through this board onto a white surface, Helmholtz was able to determine the effect of mixing two homogeneous hues of varying quantities. He discovered that every color indeed has a complement, but that the quantity of colored light required to produce white when mixing complements varies systematically with position around the color "circle."

How should this result be interpreted? Helmholtz preserved the idea that saturation, or intensity, as measured by distance from white, was a meaningful quantity. Rather than simply stipulate that homogeneous colors exhibit the maximum value of this quantity, however, he proposed to identify it with (the

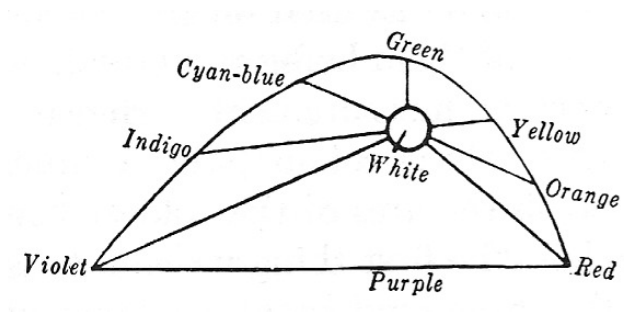

Figure 2: Helmholtz's color surface (Helmholtz, 1866, p. 139)

inverse of) the quantity of light required to produce white when mixed with a fixed quantity of monochromatic light of the complementary color. If less of a particular light is required to produce white, then, on this definition, the corresponding color is more saturated. This gave Helmholtz a means of measuring the (relative) distance from white to the edge of the color surface as a function of wavelength and, correspondingly, a new analysis of the gross structure of color space (Figure 2).

Here at last we see the two spaces which motivated Helmholtz's turn toward non-Euclidean geometry. For the points within Newton's circle and Helmholtz's truncated hyperbola are the same, i.e. the possible colors which a human can experience. Thus, if distances are treated as uniform (e.g. Euclidean) on one of these surfaces, *they cannot be uniform on the other.* How did Helmholtz himself make sense of the difference?

> As to the units of luminosity of light of different colours, in those cases where the boundary of the colour chart has been designed in the form of a circle, complementary amounts of the complementary colours, that is, amounts which produce white when they are mixed, must be considered as being equal, because by hypothesis their mixed colour, white, is midway between them....
>
> On the other hand, supposing quantities of light of different colours are to be considered as being equal, when for a certain absolute intensity of light they look equally bright to the eye, we shall get an entirely different form for the curve on which the simple colours lie, like that shown in [Figure 2]. Saturated violet and red must be farther from white than their less saturated complementary colours, because, as the eye estimates it, it takes less violet than yellow-green when we mix these two complementary colours together to get white. Hence, if white is to be in the position of their centre of gravity, the

> smaller amount of violet must have a longer lever-arm than the larger amount of yellow-green. (Helmholtz, 1866, pp. 139–40)

Helmholtz makes it clear that he considers the circular structure of the Newton color surface to be arbitrary, or at least stipulated rather than discovered. Even if this structure has some physical motivation, it does not capture the phenomenal distances between perceived colors. Conversely, the truncated hyperbola of Figure 2 represents distances in experience. In particular, by taking equal quantities of light to measure equal quantities of the respective color, we can discover inequalities in their respective saturations (where saturation is still defined as distance from white). We assume that when white is produced by mixing two complementary colors they are at equal brightness; consequently, if different quantities of these colors are required to produce white, they lie at different distances from the center point, white.

How does this fare with respect to the criteria for measurement of an arbitrary geometric space as described by Riemann? Helmholtz has used the radial structure of color space as a standard for comparison of sorts. Since the center point, white, has a distinguished position, he can use relative distance from white to determine relative distance from the center point. However, he crucially makes use of a proxy here: quantity of light (measured by width of the aperture required to make a complementary match in the experiment) is taken to stand in for (inverse) perceptual distance. If perceived brightness does not vary linearly with quantity of light, then the radial distances in Helmholtz's color space will not actually be meaningful.

Even if we grant Helmholtz success in measuring radial distance, i.e. the distance of each hue from white, it should be clear that he is still far from placing a metric on perceptual color space. In particular, what about distances between hues? In Figure 2, the distance around the circumference between indigo and violet is almost as great as that between green and orange, yet the *psychological distance* (i.e. degree of perceived dissimilarity) in the latter case is clearly far greater than in the former. An even more problematic region is that between violet and red—how should the great distance here in Helmholtz's space be interpreted? Clearly, distances around the circumference of Figure 2 are not psychologically meaningful.

Nevertheless, there is at least one great achievement of Helmholtz's empirical work here for our understanding of perceptual color space: he has succeeded in demonstrating an asymmetry in our experience of color. This also was Helmholtz's assessment of the most important contribution of this work, as he devotes significantly more discussion to the importance of asymmetries of saturation across colors than to the geometrical significance of Figure 2 in his writings.[8] Although it is only a rough ordering, his summary of this result (Helmholtz, 1866, p. 127) is borne out in modern theories of perceptual color space.

$$violet > indigo - blue > \begin{bmatrix} red \\ cyan - blue \end{bmatrix} > \begin{bmatrix} orange \\ green \end{bmatrix} > yellow$$

The greater-than symbol here indicates greater saturation. In modern color spaces, this two dimensional asymmetry is realized in three dimensions since "saturation" understood as "distance from white" is captured by two parameters: *lightness* and *chroma*, where the former is distance from perceived white (i.e. relative brightness) and the latter is degree of perceived chromaticity ("colorfulness") independent of brightness (Wyszecki and Stiles, 1982, p. 487).

4 Doing Without Just Noticeable Difference

Helmholtz's experimental result was a turning point in the application of psychophysical methods to color experience. If two complementary colors are taken as falling on "opposite" sides of color space, the experiences between them can be treated as falling on a "straight line" through that space. Mathematically, this means that variants of Fechner's measurement formula can be assumed and used to generate predictions about the metric structure of color space.[9] Experimentally, these lines suggest a dimension of comparison along which to apply Weber's method of stimulus comparison. As it turns out, the complexity of color space defeats the straightforward search for just noticeable differences, motivating the need for new experimental techniques. The purpose of this section is to explore the quality of evidence about color experience provided by variations and weakenings of the psychophysical methods for linear perceptual spaces—just how much can these methods tell us about the invariances in color experience across different individuals?

The most serious attempt to measure (something like) just noticeable differences across color space as a whole is the work of Wright (1941). Of course, by the time Wright ran these experiments, much had changed in color science from the time of Helmholtz. Of particular importance was the establishment of

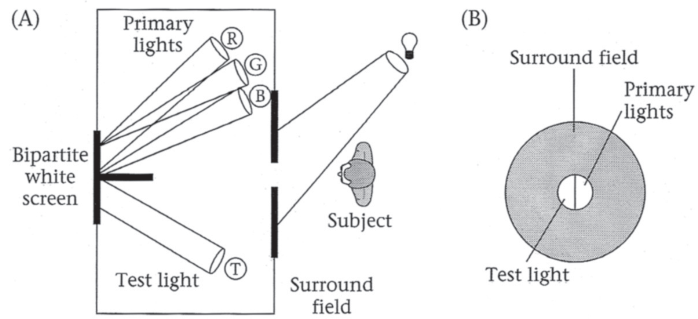

Figure 3: Setup for a color matching experiment. On the left, the view from above. On the right, the stimulus as seen by the subject. (Wandell, 1995, p. 81)

a standard color space in 1931 by the *Commission internationale de l'éclairage*, based largely on data collected earlier by Wright himself (see Fairman et al., 1997 and Brill, 1998). Although it fails to represent psychological distances accurately, the CIE 1931 color space continues to be of great importance today for the organization and reporting of color data. It is in this space that Wright made his measurements of noticeable differences.

The basic method for measuring color experiences is still in terms of comparisons between stimuli. Helmholtz's ability to isolate homogeneous stimuli from sunlight had been limited by the spectral makeup of the sunlight itself and the precision with which his machined slits could isolate narrow bands from it. Once technological developments allowed the precise measurement and control of artificial lights, a more thorough and controllable experimental setup was made possible, one not limited to bands of (roughly) homogeneous light. A subject looks through a narrow opening (width of the opening affects results) onto a bipartite white field. In the classic color matching experiment (Figure 3), a test light which can be arbitrarily complex in its wavelength composition, is shone onto one half of the bipartite field. On the other are three "primary" lights. The subject is able to adjust the intensity of each of these lights independently, and does so until the two sides of the field are perceptually identical. As it turns out, the exact composition of the primary lights does not matter (e.g. they need not correspond to "the" primary colors "red," "green," and "blue"), so long as they satisfy certain independence assumptions (i.e. are sufficiently "different"). If these independence assumptions are satisfied, subjects can perform *exact* matches with any given test light whatsoever.

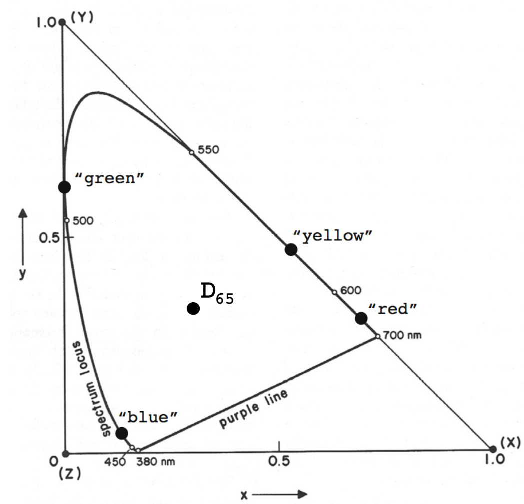

Figure 4: The CIE 1931 color space; some single-wavelength colors have been indicated around the edge of the surface. Note that no point in the interior corresponds to subjective "white" simpliciter, since assessments of white are context dependent. However, D_{65} is a CIE standard illuminant which closely approximates ordinary "white" daylight, adopted later on the recommendation of Judd et al., 1964.

Unfortunately, exact matches in a color matching experiment performed with real lights sometimes require "subtraction" of a primary. Experimentally, this can be achieved by adding that primary to the opposite side of the bipartite field. The CIE 1931 color space first averaged together results from a number of subjects on this task to form the CIE 1931 "standard observer," then linearly transformed this space into one involving three hypothetical primaries, X, Y, Z. Although these primaries (basis vectors for the space) are not actually physically realizable, they allow the representation of all possible colors via purely positive linear combinations. This space is usually represented in the (x, y) plane, where $x = \frac{X}{X+Y+Z}$ and $y = \frac{Y}{X+Y+Z}$ (Figure 4).

Even before the CIE 1931 color space was adopted, it was recognized that distances within it do not correspond to perceived differences between colors. Nevertheless, any attempt during this period to define a psychologically plausible color space was defeated by a lack of uniform data. For example, Judd (1932, 1935) emphasizes the unsatisfactory nature of data on this problem (often due to a lack of sufficient precision in the specification of the stimuli or the experimental setup). Despite this difficulty, Judd suggested a transformation of preexisting color space into one of greater psychological homogeneity which later formed the basis for the 1976 CIELUV space, recommended by the *Commission internationale de l'éclairage* as a standard for representing color differences amongst lights. Although CIELUV represents a dramatic improvement over the CIE 1931 space, it nevertheless is far from satisfactory as a psychologically plausible model of color experience (see discussion in Fairchild, 2005, pp. 194–5).

In response to the worries expressed by Judd and the *Commission internationale de l'éclairage* as a whole, Wright set out to collect a sufficiently precise, uniform data set of just noticeable differences in color space. In particular, his goal was to traverse a number of straight lines through the CIE 1931 color space, measuring just noticeable differences along the way. He rapidly dropped the restriction to "just" noticeable differences (a point to be discussed below), collecting instead data on differences judged to be perceptually equal. Wright collected data on himself and three post-graduate students, all specialists in color science or optics.

Wright's experimental setup closely matches that of Figure 3. Instead of a test light on one side and three primaries on the other, however, each side of the field was illuminated by two monochromatic lights, of the same two wavelengths, λ_1 and λ_2. These two wavelengths defined the endpoints of a line through the CIE 1931 color space, and any mixture between the two lights would fall along it. After the lights were adjusted such that both sides of the field were subjectively identical, the subject adjusted the value of his λ_1 light only until a noticeable difference with the test side of the field was detected. This light was returned to its starting point, and then the subject adjusted it in the opposite direction, again until a noticeable difference was detected. The whole procedure was repeated three times and averages taken in order to determine the step size.

Allowing the subject to adjust only only one of the two lights at a time ensured that directionality along the line of interest through the CIE 1931 color space was controlled. The intervals reported were essentially twice the noticeable difference, with the test light serving as the interval's mean. The data for Wright himself can be seen plotted within the CIE 1931 color space in Figure 5.

Why did Wright abandon *just* noticeable differences in favor of *small* noticeable differences, judged to be psychologically equivalent? The decision to rely on this "less exacting" measure than the just noticeable difference was motivated by the extraordinary number of measurements that needed to be made.

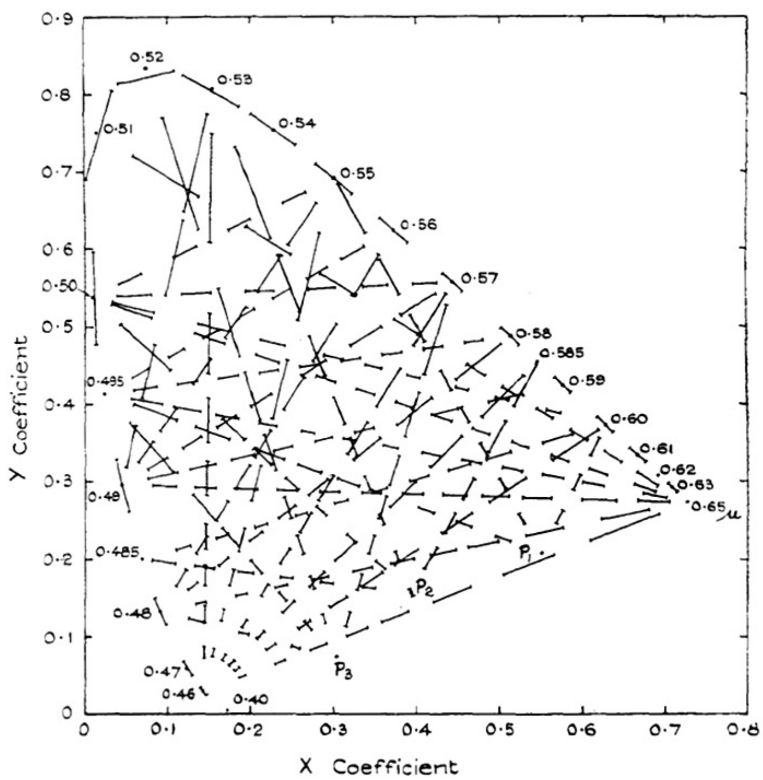

Figure 5: Equal perceptual differences graphed into the CIE 1931 color space. Each bar indicates two noticeable differences, as measured on either side of a mean color stimulus by Wright. Noticeable differences were assessed to be equal for each measurement. As such, the changes in bar length here indicate the extent to which the CIE 1931 color space distorts psychologically uniform distances in perceptual color space. (Wright, 1941, p. 99)

> It was decided that in making a step, the criterion used should be rather larger than is usually understood by a just noticeable step. In recording the latter, something akin to a sporting instinct can be introduced in an attempt to discern as minute a difference as possible, but this takes time and imposes a considerable strain on the observer. If the standard is made less exacting, the strain is reduced and a lengthy investigation can be completed more successfully. On the other hand, it is probably less easy to maintain a consistent size for the step; in the case of the author, for example, the steps recorded in the first two or three weeks of the investigation were appreciably smaller than in the later months. In time, a reasonably settled judgment was attained, and all the observations reported here were recorded after this had been achieved. (Wright, 1941, p. 96)

Wright's response to Riemann's demand for a standard which can be transported across the domain depends upon a subjective assessment of difference size. As such, it appears to be a more suspect form of measurement than the *just* noticeable difference (though see discussion by MacAdam below). In order to ensure these measurements were up to Fechnerian standards of uniformity, Wright introduced several checks on the constancy of psychological step size. One check was to return to the same lines several weeks later and retest select interval sizes. If agreement between the check and the original data was not good, the entire line would be remeasured. A second check involved comparison of Wright's data with that of his post-graduate students (who unfortunately only collected data on some of the lines which Wright himself tested). Although Wright and his students did not each choose exactly the same step size to report, after normalization, there was substantial agreement in the general pattern of changes in step size across color space.

These considerations indicate why the experiment has never been repeated. The effort and time commitment are enormous. Furthermore, individual differences ensure that there is no easy way to integrate partial data sets gathered under different experimental conditions.

> [I]t would not be very satisfactory to determine further isolated sets of results and fit them piecemeal into existing data; rather, the whole field should be tested afresh by one group of observers under one set of observing conditions. (Wright, 1941, p. 94)

Nevertheless, a number of variations on this basic psychophysical method have been used to explore the metric of subjective color space. MacAdam (1942), for example, proposed *calculating* just noticeable difference from the standard deviation in repeated color matching experiments. His motivation was dissatisfaction with the consistency provided by attempts to measure just noticeable difference directly. Interestingly, he found greater consistency when using noticeable differences of the sort employed by Wright, though the tedium of such measurements was overwhelming.

> Attempts were made to adjust color differences to equal certain standard color differences, several times greater than just noticeable.... This criterion provided more consistent results than the criterion of just noticeability, but was exceedingly tedious and required extensive training of the observer. (MacAdam, 1942, p. 257)

MacAdam's solution was simply to perform the same set of color matching experiments over and over again (approximately 50 times each) with a sophisticated measurement apparatus. This allowed him to make very precise measurements of the setting used by the subject in each match, and from that to calculate the standard deviation. The result was a sequence of ellipses when plotted into the CIE 1931 color space, each of which represented a region of perceptual indifference in a color matching experiment. The orientation and size of the ellipses revealed many of the same distortions as Wright's measurements. Unlike Wright and Judd, MacAdam argued that no simple transformation of CIE 1931 space into one with a perceptually meaningful distance metric was possible,[10] comparing CIE 1931 space to a plane map of the earth's surface, which is useful despite distortions. Although MacAdam's ellipses continue to be important for color science, they are currently believed to be a consequence of lower level physiological processes than those which determine perceptual judgments of noticeable difference (Kuehni, 2003, p. 228).

Another strategy was implemented by Wyszecki and Fielder (1971), who used a color matching setup with a tripartite viewing field. Two test lights (i) and (j) were given, and the subject was asked to adjust the primaries governing the third part of the field such that the resulting color (i, j) was a perceptual distance from each of (i) and (j) equivalent to the distance between (i) and (j). This covered the CIE 1931 color space with "perceptual equilateral triangles," which, needless to say, do not appear equilateral when plotted within that space. Again, as might be expected, the general pattern of distortions agrees with that measured by Wright and Judd, though the level of detail in the results of Wyszecki and Fielder is significantly greater.

Variations such as these on the basic color matching experimental paradigm have allowed a surprisingly precise (though still incomplete) mapping of the metric of perceptual color space. Just as in Weber's original experiments, uniform distances in the physical stimuli correspond to irregular distances in subjective experience, and the qualitative differences in these discrepancies for different parts of the spectrum motivate an asymmetrical model of color experience. But how certain can we be that these same asymmetries appear in all human visual systems? Subjects do not perform identically on any of the experiments reported above. Even in the complementary color experiment reported by Helmholtz, variations in the exact wavelength and quantity of light needed to produce white can vary across observers. Helmholtz himself reported some of these discrepancies in the second edition of the *Handbuch*. In a table of comparisons (Helmholtz, 1866, p. 128), it can be seen that the discrepancies are relatively small, less than 10 nm even at their greatest.

In the case of Wright (1941), the agreement between his own measurements and those of his students was remarkable. For most lines through the CIE 1931 space, the plots of Δl (change in step size) against l (position on the line being measured) are parallel and fall very close together. Parallelism here indicates that changes in psychological distance vary in the same direction with changes in the stimulus, while closeness indicates agreement between stimulus values at which the change occurs. The greatest amount of divergence can be found in the lower right quandrant of CIE 1931 space, where occasionally Δl grew for some subjects while shrinking for others. Another area of disagreement was in the relative size of noticeable differences as measured around the circumference of the CIE 1931 space (Figure 6). Nevertheless, these fluctuations seem minor compared to the large scale agreement about the pattern of distortion in CIE 1931 space.

MacAdam (1942) only gathered data from two subjects. His data show the same pattern of deviation as in Wright (1941); the general shape of curves graphing size of standard deviation against changes in stimulus wavelength is essentially the same for both subjects, though there are minor discrepancies. We may draw the same conclusion as with Wright's data: although there are differences in the exact internal structure of subjective color space, the overall pattern of its distortion compared to a physically uniform stimulus space appears to be the same across observers.

Wyszecki and Fielder (1971) include the most thorough discussion of variance in their data. Their experiment involved three subjects. In order to judge consistency in performance on their task, they repeated the experiment and calculated MacAdam ellipses for the third vertex of each "perceptual equilat-

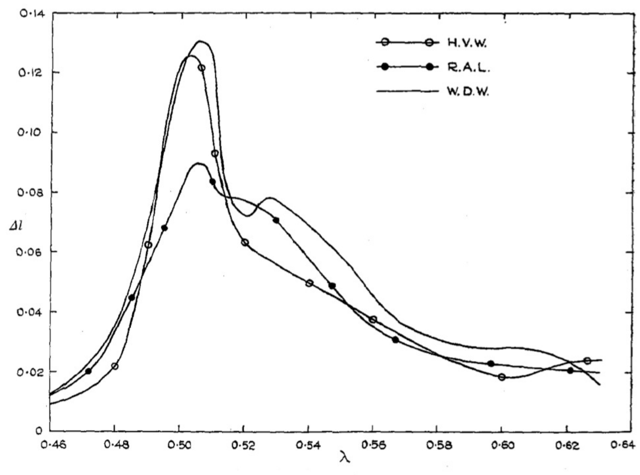

Figure 6: The degree of agreement on change in step size (Δl) as a function of wavelength in the data collected by Wright and his students (each line represents a different subject). The x-axis represents position on the circumference of the CIE 1931 color space. (Wright, 1941, p. 103)

eral triangle." Before testing variance across subjects, however, they examined consistency in the performance of a single subject on the same match performed on different days. They were surprised at how much variance they found, more than they had expected from a prior statistical analysis.

> We must accept the fact that visual color-difference-matching data obtained by the same observer on different occasions can be considerably less precisely predicted than the statistical estimates would indicate. The functioning of the visual mechanism appears to be affected by parameters that depend upon time and possibly other events and circumstances not accounted for in the model underlying statistical inferences (Wyszecki and Fielder, 1971, p. 1510)

As expected, agreement for the same subject was still greater than across subjects. When comparing the results of different subjects, agreement on what constitutes a perceptual equilateral triangle occurs about 50% of the time, within the subject's margin of error. This is determined by looking at the degree of overlap between MacAdam ellipses across subjects. Only in very few cases was there no overlap in the MacAdam ellipses for two subjects when performing the same triangular match. Although size and shape of ellipses did not seem correlated across subjects, the orientation of ellipses was strongly correlated (Wyszecki and Fielder, 1971, pp. 1510–3).

This section has surveyed some refinements in experimental techniques for measuring distances through perceptual color space. A concurrent development occurred in mathematical techniques for manipulating and analyzing psychophysical data. These range from suggested replacements of Fechner's measurement formula (e.g. Stevens, 1957) to computational techniques for extracting metric information from similarity judgments (for instance, via multidimensional scaling, as in Shepard, 1962a,b). Ultimately, the success or failure of these mathematical models rests upon the empirical support of experiments such as those discussed here.

5 Conclusion: *Do you see what I see?*

What *objective* conclusions can psychophysical methods support about the *subjective* experiences of different observers? The insight that judgments about the comparative features of two experiences for a single subject can themselves be compared across subjects makes *the relationship between subjective experiences* available as an object of scientific study. While psychophysical methods for quantifying these comparisons do not legitimate any conclusions about the *intrinsic* features of particular subjective experiences, they do legitimate conclusions about two distinct types of *extrinsic* features of experience: 1. the *structural* relationship between possible experiences; 2. the *causal* relationship between these experiences and the world.

The structural relationship between subjective experiences was initially held to be determinable *a priori*. Weber's investigation of heaviness, for example, *presupposes* that the intensity of experiences of heaviness is a single parameter which organizes those experiences into a linear order. This assumption allows him to explore systematic features of the causal relationship between experiences of heaviness and weights in the world. This methodology supported both qualitative conclusions about this relationship (e.g. that temperature causally affects judgments of heaviness) and quantitative conclusions (as captured by Weber's Law).

The more complex example of color experience, however, motivated the idea that the structural relationship between experiences needs to be investigated empirically. The crucial conceptual step required to get this program off the ground was a clear distinction between topological and metrical structure. Mathematically, this distinction is articulated clearly by Riemann. Working independently, Helmholtz discovers the same distinction in his attempts to reconcile topological conclusions derived from the continuity of color experience by Newton and Grassmann with empirical data on color combinations.

With this distinction in hand, we can see that Weber's just noticeable difference is a metrical concept; it provides a measure of distance between subjective experiences. Natural generalizations of Weber's concept turn out to be more suitable for the investigation of color experience. As we have seen, MacAdam investigated regions of indistinguishability in color experience, and Wyszecki and Fielder used the distance between two color stimuli as a standard by which to measure triples of equidistant points in color space. In each of these examples, however, some manifestation of Riemann's basic insight about metric structure can be found: a standard of comparison which can be applied in different regions of the space is required to measure distances. Weber's least distinguishable distance, MacAdam's greatest indistinguishable distance, and the subjectively equal distances of Wright and Wyszecki and Fielder all instantiate this basic idea.

Since psychophysical experiments measure the structural relationship between psychological possibilities, they can be used to investigate similarities in this structure across subjects. Additionally, psychophysical methodology presupposes a causal relationship between stimuli and experience, and therefore indirectly investigates the similarity in this causal relationship across subjects. Consequently, the greater the agreement in subjects' performance on psychophysical tasks, the greater the similarity between their respective experiences with respect to extrinsic structural and causal features. In the case of color, subjects presented with the same sets of stimuli at uniform physical distances report the same patterns of distortion in the psychological distances between their experiences. This broad agreement on the asymmetries in color experience motivates the conclusion that the structure of possible experiences of color for different subjects is the same up to a very high degree of approximation.

So, in what sense is your experience of color and mine the same? In the sense that your possible experiences of color, as identified by proxy with the stimuli which cause them, stand in the same metrical and topological relationship to each other as my experiences of color, similarly identified. Although the intrinsic features of your color experience and mine may conceivably differ when viewing the same stimulus, when considered in terms of their extrinsic relationships to each other and to the world, our experiences of color are essentially the same.

Notes

[1]Principles similar to Weber's Law as described here had been proposed before for specific types of sensation (for example, by the astronomer Pierre Bouguer (1698–1758) in the case of brightness). Weber is considered to be the first to propose it as a general principle governing many distinct sensory modalities, however.

[2]Fechner also employs this distinction. Note that Kant and Fechner's usage of "intensive" and "extensive" here differs from modern usage in physics and measurement theory.

[3]Thomas Young (1773–1829) developed his own triangular representation of color, replacing Mayer's three primary colors with red, green, and violet. He posited the existence of three fibers in the retina which vibrate sympathetically with light of the corresponding wavelengths, a theory which was later refined by Helmholtz. The Young-Helmholtz theory has been confirmed by experiments on the neurophysiology of the retina, though it is not sufficient to fully account for the phenomenology of color.

[4]In the following quote from Riemann, 1868, the term "specializations" (for *Bestimmungsweisen*) has been replaced with "determinations," following the corrections and discussion of Ferreirós, 1999, p. 63.

[5]The interaction between Helmholtz and Grassmann on color mixtures is discussed in detail by Lenoir (2006), from which this section has greatly benefited.

[6]At issue here is the mixing of colored lights, not of pigments, though it would actually be in this very series of experiments that Helmholtz himself would first clearly articulate the distinction between lights and pigments when it comes to color mixtures. Previously, the two had frequently been confused or conflated, even in the experiments of Newton.

[7]A point noticed at the time, see, for example, note III appended to Maxwell, 1857.

[8]For Helmholtz, these asymmetries provide crucial evidence for determining the wavelengths to which the three color receptors in the retina are sensitive (c.f. endnote 3). We set this issue aside here as our focus is on the phenomenology of color experience rather than the physiology of color vision.

[9] Helmholtz (1892) provides an analysis of the "shortest lines" through color space, assuming a weakened version of Fechner's measurement formula. This work was critiqued and expanded by Silberstein (1938), who demonstrated that "the whole domain of colors metricized according to Fechner's law cannot be represented without distortion by a Euclidean three-dimensional model" (82). In 1943, Silberstein compared his mathematical results with the data of MacAdam (1942) (discussed below), which he claims experimentally establish that the color surface can be represented by a differential quadratic equation, just as Helmholtz (1892) had hypothesized. This claim should be tempered somewhat by the concerns about the exact psychological meaning of MacAdam's difference ellipses, see discussion later in this section.

[10]MacAdam's assessment turned out to be correct. Already, Silberstein (1938, 1943) had demonstrated that the basic assumptions of psychophysics imply that a surface of constant brightness will have variable curvature, a result confirmed by the experiments just discussed. This curved two dimensional surface can be approximated in Euclidean three space, however, leaving open the possibility that a Euclidean color solid might approximate perceptual distances. The Optical Society of America established a committee in 1947 to investigate the possibility of defining a color solid which would be perceptually isotropic, i.e. equal distances in any direction would represent equal degrees of perceptual difference. Eventually, the com-

mittee experimentally determined that the color solid is locally Riemannian. This is due to an effect dubbed by Judd (1970) the "hue super-importance" effect. Because we are more sensitive to changes in hue than in degree of chromaticity ("colorfulness"), any color space in which hue and chroma are represented by units of equal size must have a circumference approximately twice 360°. Judd himself used the image of a crinkled fan to try and describe this effect. Although the committee eventually "forced the data into an euclidean form" (MacAdam, 1985, p. 167) in order to produce the OSA Uniform Color Scales, strictly speaking no close approximation of perceptually equal color distances in three Euclidean dimensions is possible (Kuehni and Schwarz, 2008, p. 165–167).

References

Brill, M. H. (1998). How the CIE 1931 color-matching functions were derived from Wright-Guild data (erratum). *Color Research and Application*, 23(4):259.

Ehm, W. (2010). Broad views of the philosophy of nature: Riemann, Herbart, and the "matter of the mind". *Philosophical Psychology*, 23(2):141–162.

Fairchild, M. D. (2005). *Color Appearance Models*. John Wiley & Sons, Ltd., Chichester, West Sussex, 2 edition.

Fairman, H. S., Brill, M. H., and Hemmendinger, H. (1997). How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Research and Application*, 22(1):11–23.

Fechner, G. T. (1965 [1860]b). Elements of Psychophysics, vol. 2, ch. XVI. In Herrnstein, R. J. and Boring, E. G., editors, *A Source Book in the History of Psychology*, pages 66–75. Harvard University Press.

Fechner, G. T. (1966 [1860]a). *Elements of Psychophysics*. Holt, Rinehart and Winston, New York, NY.

Ferreirós, J. (1999). *The Labyrinth of Thought: A History of Set Theory and its Role in Modern Mathematics*. Birkhäuser Verlag, Boston.

Grassmann, H. (1995 [1853]). On the theory of compound colors. In *A New Branch of Mathematics*, pages 435–49. Open Court, Chicago, IL.

Hatfield, G. (1990). *The Natural and the Normative*. MIT Press, Cambridge, MA.

Helmholtz, H. (1892). Kürzeste Linien im Farbensystem. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 3:108–122. excerpt from a memoir of the same title in *Sitzgsber. der Akademie zu Berlin*, Dec. 17, 1891.

Helmholtz, H. (1962 [1866]). *Helmholtz's Treatise on Physiological Optics*, volume 2. Dover Publications, Inc., New York, NY, 3rd edition.

Helmholtz, H. (1996 [1876]). The origin and meaning of geometrical axioms. In Ewald, W. B., editor, *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, volume 2, pages 663–689. Clarendon Press, Oxford.

Herbart, J. F. (1894 [1834]). *A Textbook in Psychology.* D. Appleton and Company, New York, NY.

Judd, D. B. (1932). Chromaticity sensibility to stimulus differences. *Journal of the Optical Society of America*, 22:72–108.

Judd, D. B. (1935). A Maxwell triangle yielding uniform chromaticity scales. *Journal of the Optical Society of America*, 25:24–35.

Judd, D. B. (1970). Ideal color space. *Color Engineering*, 8(2):37–52.

Judd, D. B., MacAdam, D. L., and Wyszecki, G. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, 54(8):1031–1040.

Kant, I. (1998 [1781]). *Critique of Pure Reason.* Cambridge UP, New York, NY.

Kuehni, R. G. (2003). *Color Space and Its Divisions: Color Order from Antiquity to the Present.* John Wiley & Sons, Inc., New York, NY.

Kuehni, R. G. and Schwarz, A. (2008). *Color Ordered: A Survey of Color Order Systems from Antiquity to the Present.* Oxford UP, New York, NY.

Lenoir, T. (2006). Operationalizing Kant: Manifolds, models, and mathematics in Helmholtz's theories of perception. In Friedman, M. and Nordmann, A., editors, *The Kantian Legacy in Nineteenth-Century Science*, pages 141–210. MIT Press, Cambridge, MA.

MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America*, 32:247–274.

MacAdam, D. L. (1985). *Color Measurement: Theme and Variations*. Springer-Verlag, Berlin, 2 edition.

Maxwell, J. C. (1857). Experiments on colour, as perceived by the eye, with remarks on colour-blindness. *Transactions of the Royal Society of Edinburgh*, XXI:275–298.

Newton, I. (1952 [1730]). *Opticks, or A Treatise of the Reflections, Refractions, Inflections & Colours of Light*. Dover Publications, Inc., New York, NY, fourth edition.

Riemann, G. F. B. (1996 [1868]). On the hypotheses which lie at the foundation of geometry. In Ewald, W. B., editor, *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, volume 2, pages 652–661. Clarendon Press, Oxford.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2):125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. ii. *Psychometrika*, 27(3):219–246.

Silberstein, L. (1938). Investigations on the intrinsic properties of the color domain. *Journal of the Optical Society of America*, 28:63–85.

Silberstein, L. (1943). Investigations on the intrinsic properties of the color domain. II. *Journal of the Optical Society of America*, 33(1):1–10.

Stevens, S. S. (1957). On the psychophysical law. *The Psychological Review*, 64(3):153–181.

Wandell, B. A. (1995). *Foundations of Vision*. Sinauer, Sunderland, MA.

Weber, E. H. (1978 [1846]). Der Tastsinn und das Gemeingefühl. In *E. H. Weber: The Sense of Touch*, pages 137–268. Academic Press, New York, NY.

Wright, W. D. (1941). The sensitivity of the eye to small colour differences. *Proc. Phys. Soc.*, 53(296):93–112.

Wyszecki, G. and Fielder, G. H. (1971). Color-difference matches. *Journal of the Optical Society of America*, 61:1501–1513.

Wyszecki, G. and Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley & Sons, Inc., New York, NY, 2nd edition.